Library Progress International *Print version ISSN 0970 1052*
Vol.44 No.3, Jul-Dec 2024: P.20509-20518 *Online version ISSN 2320 317X*

Original Article *Available online at www.bpasjournals.com*

A New Efficient Estimator of Finite Population Mean for Dealing Non-Response in Survey Sampling

[1]Udita Gupta,[2]Prof. Manoj Kumar Srivastava,[3]Prof. Namita Srivastava,[4]Anjali Bhardwaj

[1] drudita0108@gmail.com Research Scholar Department of Statistics, I.S.S., Dr. Bhimrao Ambedkar University,Agra
[2] mksiss87@gmail.com Professor Institute of Social Sciences, Dr. Bhimrao Ambedkar University, Agra
[3] drnamita.sjc@gmail.com Professor and Head of Department of Statistics St. John's College, Agra
[4]anjalibharadwaj32@gmail.com Research Scholar Department of Statistics, I.S.S.,Dr. Bhimrao Ambedkar University, Agra

How to cite this article: Udita Gupta, Manoj Kumar Srivastava, Namita Srivastava,Anjali Bhardwaj (2024) A New Efficient Estimator of Finite Population Mean for Dealing Non-Response in Survey Sampling. *Library Progress International,* 44(3), 20509-20518.

Abstract
This paper, we propose a new exponential product ratio estimator and an exponential regression estimator to estimate the mean value $\bar{Y}$ of a finite population of the study variable y using the auxiliary variable for dealing non-response. Using a large sample approximation, the expression of bias and mean error squared (MSE) are obtained for the proposed estimates. Under certain realistic conditions, the proposed estimation has been found to be more or equal accurate than normal unbiased estimates. The effectiveness of the proposed estimate is demonstrated by empirical studies in comparison with sample mean, ratio estimators and other current estimates.

Keywords: Auxiliary Variable; Study Variable; Bias; Mean Squared Error; Non-Response; Ratio-product exponential estimator; Regression exponential estimator; Searls Constant

1. INTRODUCTION

To improve the efficiency of the estimators, auxiliary information is generally used. Cochran (1940) proposed a ratio estimator for estimating the population mean when the study variable is positively correlated with auxiliary variables. Murthy (1964) proposed a product estimation that predicted the population mean when both study and auxiliary variables are negatively correlated, in contrast to the ratio estimator situation. Hansen et. al (1953) proposed the difference estimator which was modified to provide the linear regression estimator for the population mean or total.

Almost all surveys suffer from lack of response (non-response) in practice. If the subjects are refused, absenteeism occurs and sometimes lack of information is a reason for not responding. The pioneering work of Hansen and Hurwitz (1946), assumed that a sub-sample of initial non-respondents is recontacted with a more expensive method, suggesting the first attempt by mail questionnaire and the second attempt by a personal interview. Sampling experts are sometimes using auxiliary information in order to increase the accuracy of estimates when determining population parameters such as mean, total or ratio.

In the sampling literature, it is well known that the efficiency of the population mean estimator of study variable y can be increased by using auxiliary information that is highly correlated with study variable y. Rao (1986), Khare and Srivastava (1995, 1997), Okafor and Lee (2000) and Singh and Kumar (2008, 2009, 2010) have proposed a tool to estimate the population mean of the study variable y by using auxiliary information in the presence of non-response.

The remaining part of the paper is organized as follows: in section 2, notations under nonresponse is introduced. In section 3, existing estimators under non-response. In section 4, proposed estimator for estimating finite population mean is defined. In section 5, theoretical comparison is done between existing and proposed estimator. In section 6, existing and proposed estimators are compared numerically. Finally, section 7 condenses the principal discovery and culminate the document.

2. NOTATION

For a finite population P= $(P_1, P_2, P_3 \dots P_N)$ of size N and a random sample of size n is drawn without replacement. Let the characteristics under study, y (say) takes value y_i on the unit $P_i, (i = 1,2,3 \dots N)$. In survey on the human population, it is often the case that n_i unit respond on the first attempt while n_2 $(= n - n_1)$ units do not provide any response. In the case of non-response at the initial stage, Hansen and Hurwitz (1946) proposed a double sampling plan for estimating the population mean having the steps given below:

a) A simple random sample (SRS) of size n is drawn and the questionnaire is mailed to the sample units

b) A sub-sample of size $q = (n_2/h), (h \geq 1)$ from the n_2 non-responding units in the initial attempt is contacted through personal interviews.

In the Hansen and Hurwitz method the population is supposed to be consisting of response stratum of size N_1 and the non-response stratum of size $N_2 = (N - N_1)$. Let $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$ and $S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})^2$ denote the population mean and variance of the study variable y. Let $\bar{Y}_1 = \frac{1}{N_1}\sum_{i=1}^{N_1} y_i$ and $S_{y(1)}^2 = \frac{1}{N_1-1}\sum_{i=1}^{N_1}(y_i - \bar{Y}_1)^2$ denote the mean and variance of the respondent group (or strata). Similarly, let $\bar{Y}_2 = \frac{1}{N_2}\sum_{i=1}^{N_2} y_i$ and $S_{y(2)}^2 = \frac{1}{N_2-1}\sum_{i=1}^{N_2}(y_i - \bar{Y}_2)^2$ denote the mean and variance of the non-respondent group (or strata). The population mean can be written as

$$\bar{Y} = W_1\bar{Y}_1 + W_2\bar{Y}_2$$

where $W_1 = \left(\frac{N_1}{N}\right)$ and $W_2 = \left(\frac{N_2}{N}\right)$. The sample mean $\bar{y}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} y_i$ denote the mean of the n_1 responding units and $\bar{y}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} y_i$ denote the mean of the n_2 non-responding units.

Let $\bar{y}_{2q} = \frac{1}{q}\sum_{i=1}^{q} y_i$ denote the mean of the q sub-sampled units where q= $\frac{n_2}{h}$. Hansen and Hurwitz (1946) suggested an unbiased estimator for the population mean $\bar{Y}$ of the study variable y is given as:

$$\bar{y}^* = w_1\bar{y}_1 + w_2\bar{y}_{2\mathrm{q}} \tag{2.1}$$

where $w_1 = \frac{n_1}{n}$ and $w_2 = \frac{n_2}{n}$ are responding proportions and non-responding proportions of the sample. The variance $\bar{y}^*$ is given below:

$$V(\bar{y}^*) = \bar{Y}^2\left[\left(\frac{1-f}{n}\right)C_y^2 + \frac{W_2(h-1)}{n}C_{y(2)}^2\right] \tag{2.2}$$

where $C_y^2 = \frac{S_y^2}{\bar{Y}^2}$ and $C_{y(2)}^2 = \frac{S_{y(2)}^2}{\bar{Y}^2}$.

Let $x_i (i = 1,2, \dots N)$ denote the auxiliary variable correlated with the study variable $y_i (i = 1,2, \dots, N)$. Let $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} x_i$ and $S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{X})^2$ denote the population mean and variance of the auxiliary variable y. Let $\bar{X}_1 = \frac{1}{N_1}\sum_{i=1}^{N_1} x_i$ and $S_{x(1)}^2 = \frac{1}{N_1-1}\sum_{i=1}^{N_1}(x_i - \bar{X}_1)^2$ denote the mean and variance of the respondent group (or strata). Similarly, let $\bar{Y}_2 = \frac{1}{N_2}\sum_{i=1}^{N_2} y_i$ and $S_{x(2)}^2 = \frac{1}{N_2-1}\sum_{i=1}^{N_2}(x_i - \bar{X}_2)^2$ denote the mean and variance of the non-respondent group (or strata). Let $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ denote the mean of all the n units. Let $\bar{x}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} x_i$ denote the mean of the n_1 responding units and $\bar{x}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} x_i$ denote the mean of the n_2 non-responding units.

Let $\bar{x}_{2q} = \frac{1}{q}\sum_{i=1}^{q} x_i$ denote the mean of the q sub-sampled units where $q = \frac{n_2}{h}$. With this background, define an unbiased estimator of population mean $\bar{X}$ is given as:

$$\bar{x}^* = w_1\bar{x}_1 + w_2\bar{x}_{2q} \tag{2.3}$$

The variance of $\bar{x}^*$ is given below:

$$V(\bar{x}^*) = \bar{X}^2\left[\left(\frac{1-f}{n}\right)C_x^2 + \frac{W_2(h-1)}{n}C_{x(2)}^2\right] \tag{2.4}$$

where $C_x^2 = \frac{S_x^2}{\bar{X}^2}$ and $C_{x(2)}^2 = \frac{S_{x(2)}^2}{\bar{X}^2}$.

2.1 Bias and Mean Square Error (MSE) of the Proposed Estimators

Let us define the following terms:

$\bar{y}^* = \bar{Y}(1+\varepsilon_0)$

$\bar{x}^* = \bar{X}(1+\varepsilon_1)$

$$E(\varepsilon_0) = E(\varepsilon_1) = 0 \tag{2.5}$$

$$E(\varepsilon_0^2) = \left[\left(\frac{1-f}{n}\right)C_y^2 + \frac{W_2(h-1)}{n}C_{y(2)}^2\right] = B\ (say) \tag{2.6}$$

$$E(\varepsilon_1^2) = \left[\left(\frac{1-f}{n}\right)C_x^2 + \frac{W_2(h-1)}{n}C_{x(2)}^2\right] = A\ (say) \tag{2.7}$$

$$E(\varepsilon_0\varepsilon_1) = \left[\left(\frac{1-f}{n}\right)\rho_{yx}C_yC_x + \frac{W_2(h-1)}{n}\rho_{yx(2)}C_{y(2)}C_{x(2)}\right] = C\ (say) \tag{2.8}$$

where

$\rho_{yx} = S_{yx}/S_xS_y$;

$\rho_{yx(2)} = S_{yx(2)}/S_{x(2)}S_{y(2)}$

$$S_{yx} = \frac{1}{N-1}\sum_{i=1}^{N}(y_i-\bar{Y})(x_i-\bar{X})$$

$$S_{yx(2)} = \frac{1}{N_2-1}\sum_{i=1}^{N_2}(y_i-\bar{Y}_2)(x_i-\bar{X}_2)$$

3. Existing Estimator

Where only a few observations are missing from the sample. Initial, estimators for estimating population mean $\bar{Y}$ was suggested by Hansen and Hurwitz (1946). Many other authors work for the similar situation. In this case, few existing estimates of population mean $\bar{Y}$ under nonresponse are given.

(i) Rao (1986) suggested a ratio estimator for estimation the population mean $\bar{Y}$ of the study variable y is given by

$$t_r^* = \bar{y}^*\left(\frac{\bar{X}}{\bar{x}^*}\right) \tag{3.1}$$

$$B(t_r^*) = \bar{Y}(A-C)$$

$$MSE(t_r^*) = \bar{Y}^2(B+A-2C)$$

$$MSE(t_r^*) = \bar{Y}^2\left[\left(\frac{1-f}{n}\right)\{C_y^2 + C_x^2 - 2\rho_{yx}C_yC_x\} + \frac{W_2(h-1)}{n}\left[C_{y(2)}^2 + C_{x(2)}^2 - 2\rho_{yx(2)}C_{y(2)}C_{x(2)}\right]\right] \tag{3.2}$$

(ii) Khare and Srivastava (1993) suggested a product estimator for the population mean $\boldsymbol{\bar{Y}}$ of the study variable $\boldsymbol{y}$ is given as

$$t_{ks}^* = \bar{y}^*\left(\frac{\bar{x}^*}{\bar{X}}\right) \tag{3.3}$$

$$B(t_{ks}^*) = \bar{Y}C$$

$$MSE(t_{ks}^*) = \bar{Y}^2(A + B + 2C)$$

$$MSE(t_{ks}^*) = \bar{Y}^2\left[\left(\frac{1-f}{n}\right)\{C_y^2 + C_x^2 + 2\rho_{yx}C_yC_x\} + \frac{W_2(h-1)}{n}\left[C_{y(2)}^2 + C_{x(2)}^2 + 2\rho_{yx(2)}C_{y(2)}C_{x(2)}\right]\right] \quad (3.4)$$

(iii) Singh et. al. (2008) suggested an exponential ratio type estimators for the population mean $\bar{Y}$ of the study variable y are

$$t_{er}^* = \bar{y}^* exp\left[\frac{\bar{X} - \bar{x}^*}{\bar{X} + \bar{x}^*}\right] \quad (3.5)$$

$$B(t_{er}^*) = \frac{\bar{Y}}{8}(3A - 4C)$$

$$MSE(t_{er}^*) = \bar{Y}^2\left(B + \frac{A}{4} - C\right)$$

$$MSE(t_{er}^*) = \bar{Y}^2\left[\left(\frac{1-f}{n}\right)\left\{C_y^2 + \frac{C_x^2}{4} - \rho_{yx}C_yC_x\right\} + \frac{W_2(h-1)}{n}\left[C_{y(2)}^2 + \frac{C_{x(2)}^2}{4} - \rho_{yx(2)}C_{y(2)}C_{x(2)}\right]\right] \quad (3.6)$$

(iv) Singh et. al. (2008) suggested an exponential product type estimators for the population mean $\bar{Y}$ of the study variable y are

$$t_{ep}^* = \bar{y}^* exp\left[\frac{\bar{x}^* - \bar{X}}{\bar{x}^* + \bar{X}}\right] \quad (3.7)$$

$$B(t_{ep}^*) = \frac{\bar{Y}}{8}(4C - A)$$

$$MSE(t_{ep}^*) = \bar{Y}^2\left(B + \frac{A}{4} + C\right)$$

$$MSE(t_{ep}^*) = \bar{Y}^2\left[\left(\frac{1-f}{n}\right)\left\{C_y^2 + \frac{C_x^2}{4} + \rho_{yx}C_yC_x\right\} + \frac{W_2(h-1)}{n}\left[C_{y(2)}^2 + \frac{C_{x(2)}^2}{4} + \rho_{yx(2)}C_{y(2)}C_{x(2)}\right]\right] \quad (3.8)$$

4. Proposed Estimators

Strategy I: $\bar{y}^*, \bar{x}^*$ *and* $\bar{X}$ are used. The non-response happens on both study variable y as well as auxiliary variable x, and the population mean $\bar{X}$ of the auxiliary variable is known. In this strategy, we proposed a regression exponential type estimator and the estimator is

$$t_{sh}^* = (\bar{y}^* + \beta(\bar{X} - \bar{x}^*)) \exp\left[\frac{\bar{X} - \bar{x}^*}{\bar{X} + \bar{x}^*}\right]$$

Case-1: We proposed estimator as regression exponential type estimator in which non-response occurs in study as well as auxiliary variable

$$t_{sh}^{*} = (\bar{y}^{*} + \beta(\bar{X} - \bar{x}^{*}))\exp\left[\frac{\bar{X} - \bar{x}^{*}}{\bar{X} + \bar{x}^{*}}\right] \tag{4.1}$$

where β is a real constant to be determined such that the MSE of t_{sh}^{*} is minimum.

Now, expressing t_{sh}^{*} in terms of $\varepsilon's$ we have

$$t_{sh}^{*} = [\bar{Y}(1+\varepsilon_0) + \beta\{\bar{X} - \bar{X}(1+\varepsilon_1)\}]\exp\left[\frac{\bar{X} - \bar{X}(1+\varepsilon_1)}{\bar{X} + \bar{X}(1+\varepsilon_1)}\right]$$

$$= [\bar{Y}(1+\varepsilon_0) + \beta\bar{X}(-\varepsilon_1)]\exp\left[\frac{-\varepsilon_1}{2+\varepsilon_1}\right]$$

$$= [\bar{Y}(1+\varepsilon_0) - \beta\bar{X}\varepsilon_1]exp\left(-\frac{\varepsilon_1}{2}\right)\left(1+\frac{\varepsilon_1}{2}\right)^{-1} \tag{4.2}$$

Expanding the right-hand side of equation (4.2) and neglecting the terms involving powers of $\varepsilon's$ greater than two, we have

$$t_{sh}^{*} - \bar{Y} = \bar{Y}\left[\varepsilon_0 - \frac{\varepsilon_1}{2} - \frac{\varepsilon_0\varepsilon_1}{2} + \frac{3}{8}\varepsilon_1^2\right] - \beta\bar{X}\left[\varepsilon_1 - \frac{\varepsilon_1^2}{2}\right] \tag{4.3}$$

Taking Expectation on both sides of (4.3), we get the bias of the estimator t_{sh}^{*} as

$$B(t_{sh}^{*}) = \frac{\bar{Y}}{8}(3A - 4C) + \frac{1}{2}\beta\bar{X}A \tag{4.4}$$

Squaring of (4.3) on both sides and neglecting terms of $\varepsilon's$ involving power greater than two, we have

$$(t_{sh}^{*} - \bar{Y})^2 = \bar{Y}^2\left[\varepsilon_0^2 + \frac{\varepsilon_1^2}{4} - \varepsilon_0\varepsilon_1\right] + \beta^2\bar{X}^2\varepsilon_1^2 - 2\beta\bar{X}\bar{Y}\left[\varepsilon_0\varepsilon_1 - \frac{\varepsilon_1^2}{2}\right] \tag{4.5}$$

Taking Expectation on both sides of (4.5), we get the exact mean square error (MSE) of t_{sh}^{*} and approximation of MSE (to the first degree of approximation) of t_{sh}^{*}, as

$$MSE(t_{sh}^{*}) = \frac{\bar{Y}^2}{4}(4B + A - 4C) + \beta^2\bar{X}^2A - \beta\bar{X}\bar{Y}(2C - A) \tag{4.6}$$

Differentiate equation (4.6) w.r.to β and equating to zero for obtaining the optimum value of β. The optimum value of β which makes the MSE minimum of equation (4.6) is given by

$$\beta = \frac{R(2C - A)}{2A} \tag{4.7}$$

where $R = \bar{Y}/\bar{X}$

Substitute the optimum value of 'β' from (4.7) in (4.6), we get the exact MSE of the optimum estimator $t_{sh(opt)}^{*}$

$$MSE\left(t_{sh(opt)}^{*}\right) = \min MSE\,(t_{sh}^{*}) = \bar{Y}^2\left[B - \frac{C^2}{A}\right] \tag{4.8}$$

$$MSE\left(t_{sh(opt)}^{*}\right) = \bar{Y}^2\left[\left\{\left(\frac{1-f}{n}\right)C_y^2 + \frac{W_2(h-1)}{n}C_{y(2)}^2\right\} - \frac{\left\{\left(\frac{1-f}{n}\right)\rho_{yx}C_yC_x + \frac{W_2(h-1)}{n}\rho_{yx(2)}C_{y(2)}C_{x(2)}\right\}^2}{\left\{\left(\frac{1-f}{n}\right)C_x^2 + \frac{W_2(h-1)}{n}C_{x(2)}^2\right\}}\right] \tag{4.8b}$$

Strategy II: Donald T. Searls (1964) introduce a constant which results in higher efficient estimators. In this strategy, we propose to use Searls (1964) type transformation (STT) under strategy I. The STT is defined as

$$T = \alpha\bar{y}^*$$

where α is a suitable chosen scalar. The following estimators under strategy I described above are proposed using the STT. The proposed estimator under strategy II, when $\bar{y}^*, \bar{x}^*$ and $\bar{X}$ are used, is given by

$$t_{sh1}^* = (\alpha\bar{y}^* + \beta(\bar{X} - \bar{x}^*)) \exp\left[\frac{\bar{X} - \bar{x}^*}{\bar{X} + \bar{x}^*}\right]$$

Case-2: We proposed estimator as regression exponential type estimator in which non-response occurs in study as well as auxiliary variable

$$t_{sh1}^* = (\alpha\bar{y}^* + \beta(\bar{X} - \bar{x}^*)) \exp\left[\frac{\bar{X} - \bar{x}^*}{\bar{X} + \bar{x}^*}\right] \quad (4.9)$$

where α and β is a real constant to be determined such that the MSE of t_{sh1}^* is minimum.

Now, expressing t_{sh1}^* in terms of $\varepsilon's$ we have

$$t_{sh1}^* = [\alpha\bar{Y}(1+\varepsilon_0) + \beta\{\bar{X} - \bar{X}(1+\varepsilon_1)\}] \exp\left[\frac{\bar{X} - \bar{X}(1+\varepsilon_1)}{\bar{X} + \bar{X}(1+\varepsilon_1)}\right]$$

$$= [\alpha\bar{Y}(1+\varepsilon_0) + \beta\bar{X}(-\varepsilon_1)] \exp\left[\frac{-\varepsilon_1}{2+\varepsilon_1}\right]$$

$$= [\alpha\bar{Y}(1+\varepsilon_0) - \beta\bar{X}\varepsilon_1] exp\left(-\frac{\varepsilon_1}{2}\right)\left(1+\frac{\varepsilon_1}{2}\right)^{-1} \quad (4.10)$$

Expanding the right-hand side of equation (4.10) and neglecting the terms involving powers of $\varepsilon's$ greater than two, we have

$$t_{sh1}^* - \bar{Y} = (\alpha - 1)\bar{Y} + \alpha\bar{Y}\left[\varepsilon_0 - \frac{\varepsilon_1}{2} - \frac{\varepsilon_0\varepsilon_1}{2} + \frac{3}{8}\varepsilon_1^2\right] - \beta\bar{X}\left[\varepsilon_1 - \frac{\varepsilon_1^2}{2}\right] \quad (4.11)$$

Taking Expectation on both sides of (4.11), we get the bias of the estimator t_{sh1}^* as

$$B(t_{sh1}^*) = (\alpha - 1)\bar{Y} + \frac{\alpha\bar{Y}}{8}(3A - 4C) + \frac{1}{2}\beta\bar{X}A \quad (4.12)$$

Squaring of (4.11) on both sides and neglecting terms of $\varepsilon's$ involving power greater than two, we have

$$(t_{sh1}^* - \bar{Y})^2 = \left[(\alpha - 1)\bar{Y} + \alpha\bar{Y}\left[\varepsilon_0 - \frac{\varepsilon_1}{2} - \frac{\varepsilon_0\varepsilon_1}{2} + \frac{3}{8}\varepsilon_1^2\right] - \beta\bar{X}\left[\varepsilon_1 - \frac{\varepsilon_1^2}{2}\right]\right]^2$$

$$= (\alpha - 1)^2\bar{Y}^2 + \alpha^2\bar{Y}^2\left[\varepsilon_0^2 + \frac{\varepsilon_1^2}{4} - \varepsilon_0\varepsilon_1\right] + \beta^2\bar{X}^2\varepsilon_1^2 + 2\alpha(\alpha - 1)\bar{Y}^2\left[\varepsilon_0 - \frac{\varepsilon_1}{2} - \frac{\varepsilon_0\varepsilon_1}{2} + \frac{3}{8}\varepsilon_1^2\right] - 2(\alpha - 1)\beta\bar{X}\bar{Y}\left[\varepsilon_1 - \frac{\varepsilon_1^2}{2}\right] - 2\alpha\beta\bar{X}\bar{Y}\left[\varepsilon_0\varepsilon_1 - \frac{\varepsilon_1^2}{2}\right] \quad (4.13)$$

Taking Expectation on both sides of (4.13), we get the exact mean square error (MSE) of t_{sh1}^* and approximation of MSE (to the first degree of approximation) of t_{sh1}^*, as

$$MSE(t_{sh1}^*) = (\alpha - 1)^2\bar{Y}^2 + \alpha^2\bar{Y}^2\left[B + \frac{A}{4} - C\right] + \beta^2\bar{X}^2A + \alpha(\alpha - 1)\bar{Y}^2\left[\frac{3}{4}A - C\right] + (\alpha - 1)\beta A\bar{X}\bar{Y} - \alpha\beta\bar{Y}\bar{X}(2C - A) \quad (4.14)$$

Differentiate equation (4.14) w.r.to α and β and equating to zero for obtaining the optimum value of α and β. The optimum value of α and β which makes the MSE minimum of equation (4.14) are given by

$$\alpha = \frac{A - \frac{A^2}{8}}{A - C^2 + AB} = V_0(say)$$

$$\beta = \frac{R}{A + AB - C^2}\left[\frac{A^2 - 4A - 4C^2 + 8C + 4AB - AC}{8}\right] = V_1(say)$$

where $R = \bar{Y}/\bar{X}$

Substitute the optimum value of 'α' and 'β' in (4.14), we get the expression of minimum MSE of the estimator $t_{sh1(opt)}^*$

$$MSE\left(t^*_{sh1(opt)}\right) = (V_0-1)^2\bar{Y}^2 + \bar{Y}^2\left[B+\frac{A}{4}-C\right] + {V_1}^2\bar{X}^2A + V_0(V_0-1)\bar{Y}^2\left[\frac{3}{4}A-C\right] + (V_0-1)V_1A\bar{X}\bar{Y} - V_0V_1\bar{Y}\bar{X}(2C-A) \qquad (4.15)$$

5. Theoretical Efficiency Comparison

The MSE's of the existing estimators to the first degree of approximation are derived as:

$$Var\,(\bar{y}^*) = \bar{Y}^2B \qquad (5.1)$$

$$MSE(t_r^*) = \bar{Y}^2(B+A-2C) \qquad (5.2)$$

$$MSE(t_{ks}^*) = \bar{Y}^2(A+B+2C) \qquad (5.3)$$

$$MSE(t_{er}^*) = \bar{Y}^2\left(B+\frac{A}{4}-C\right) \qquad (5.4)$$

$$MSE\left(t_{ep}^*\right) = \bar{Y}^2\left(B+\frac{A}{4}+C\right) \qquad (5.5)$$

Below is the comparison between proposed estimators with another existing estimator. Efficiency condition over some related existing estimators.

1) $Var\,(\bar{y}^*) - MSE\left(t^*_{sh(opt)}\right) = \bar{Y}^2\left(B-B+\frac{C^2}{A}\right) = \bar{Y}^2\frac{C^2}{A} \geq 0 \qquad (5.6a)$

$Var(\bar{y}^*) - MSE\left(t^*_{sh1(opt)}\right)\, if$

$$= \bar{Y}^2B - (V_0-1)^2\bar{Y}^2 - \bar{Y}^2\left[B+\frac{A}{4}-C\right] - {V_1}^2\bar{X}^2A - V_0(V_0-1)\bar{Y}^2\left[\frac{3}{4}A-C\right] - (V_0-1)V_1A\bar{X}\bar{Y} + V_0V_1\bar{Y}\bar{X}(2C-A) \geq 0 \qquad (5.6b)$$

1) $MSE(t_r^*) - MSE\,\left(t^*_{sh(opt)}\right) = \bar{Y}^2\left(B+A-2C-B+\frac{C^2}{A}\right)$

$$= \bar{Y}^2\left(A-2C+\frac{C^2}{A}\right)$$

$$= \bar{Y}^2\left(\frac{A^2-2CA+C^2}{A}\right) = (A-C)^2 \geq 0 \qquad (5.7a)$$

$MSE(t_r^*) - MSE\,\left(t^*_{sh1(opt)}\right)\, if$

$$= \bar{Y}^2(B+A-2C) - (V_0-1)^2\bar{Y}^2 - \bar{Y}^2\left[B+\frac{A}{4}-C\right] - {V_1}^2\bar{X}^2A - V_0(V_0-1)\bar{Y}^2\left[\frac{3}{4}A-C\right] - (V_0-1)V_1A\bar{X}\bar{Y} + V_0V_1\bar{Y}\bar{X}(2C-A) \geq 0 \qquad (5.7b)$$

2) $MSE(t_{ks}^*) - MSE\left(t^*_{sh(opt)}\right) = \bar{Y}^2\left(A+B+2C-B+\frac{C^2}{A}\right)$

$$= \bar{Y}^2\left(A+2C+\frac{C^2}{A}\right)$$

$$= \bar{Y}^2\left(\frac{A^2+2CA+C^2}{A}\right) = (A+C)^2 \geq 0 \qquad (5.8a)$$

$MSE(t_{ks}^*) - MSE\,\left(t^*_{sh1(opt)}\right)\, if$

$$= \bar{Y}^2(B+A+2C) - (V_0-1)^2\bar{Y}^2 - \bar{Y}^2\left[B+\frac{A}{4}-C\right] - {V_1}^2\bar{X}^2A - V_0(V_0-1)\bar{Y}^2\left[\frac{3}{4}A-C\right] - (V_0-1)V_1A\bar{X}\bar{Y} + V_0V_1\bar{Y}\bar{X}(2C-A) \geq 0 \qquad (5.8b)$$

3) $$MSE(t_{er}^{*}) - MSE\left(t_{sh(opt)}^{*}\right) = \bar{Y}^2\left(B + \frac{A}{4} - C - B + \frac{C^2}{A}\right)$$

$$= \bar{Y}^2\left(\frac{A}{4} - C + \frac{C^2}{A}\right)$$

$$= (A - 2C)^2 \geq 0 \qquad (5.9a)$$

$$MSE(t_{er}^{*}) - MSE\left(t_{sh1(opt)}^{*}\right) \; if$$

$$= \bar{Y}^2\left(B + \frac{A}{4} - C\right) - (V_0 - 1)^2\bar{Y}^2 - \bar{Y}^2\left[B + \frac{A}{4} - C\right] - V_1{}^2\bar{X}^2 A - V_0(V_0 - 1)\bar{Y}^2\left[\frac{3}{4}A - C\right]$$

$$- (V_0 - 1)V_1 A\bar{X}\bar{Y} + V_0 V_1 \bar{Y}\bar{X}(2C - A) \geq 0 \qquad (5.9b)$$

4) $$MSE\left(t_{ep}^{*}\right) - MSE\left(t_{sh(opt)}^{*}\right) = \bar{Y}^2\left(B + \frac{A}{4} + C - B + \frac{C^2}{A}\right)$$

$$= \bar{Y}^2\left(\frac{A}{4} + C + \frac{C^2}{A}\right)$$

$$= (A + 2C)^2 \geq 0 \qquad (5.10a)$$

$$MSE\left(t_{ep}^{*}\right) - MSE\left(t_{sh1(opt)}^{*}\right) \; if$$

$$= \bar{Y}^2\left(B + \frac{A}{4} + C\right) - (V_0 - 1)^2\bar{Y}^2 - \bar{Y}^2\left[B + \frac{A}{4} - C\right] - V_1{}^2\bar{X}^2 A - V_0(V_0 - 1)\bar{Y}^2\left[\frac{3}{4}A - C\right]$$

$$- (V_0 - 1)V_1 A\bar{X}\bar{Y} + V_0 V_1 \bar{Y}\bar{X}(2C - A) \geq 0 \qquad (5.10b)$$

Under the condition derived above proposed estimator proves to be more efficient than the existing traditional estimator. In the next section, the theoretical findings are supported by a numerical study.

6. Numerical Illustration

To illustrate numerical meaning of the theoretical results, consider a real dataset used by Khare and Srivastava (1993). The description of the dataset is given below:

A list of 70 villages in India along their population in 1981 and cultivated area (in acres) in the same year is considered (Singh and Choudhary, 1986). Here, the cultivated area (in acres) is taken as the main study variable and the population of the village is taken as the auxiliary variable. The parameters of the population are as follows:

$$\bar{Y} = 981.29, \quad \bar{X} = 1755.53, \quad C_y = 0.6254, \quad C_x = 0.8009, \quad \bar{Y}_2 = 597.29, \quad \bar{X}_2 = 1100.24,$$

$C_{y(2)} = 0.4087, \quad C_{x(2)} = 0.5739, \quad \rho_{yx} = 0.778, \quad \rho_{yx(2)} = 0.445, \quad W_2 = 0.20,$

$N = 70, \quad n = 35$

We computed the percent-relative efficiency (PRE's) of various existing estimators with respect to the usual unbiased estimator $\bar{y}^*$ for different values of h, by using the formulae

$$PRE(t^*, \bar{y}^*) = \frac{Var\ (\bar{y}^*)}{MSE\ (t^*)} \times 100$$

where $t^* = t_r^*, t_{ks}^*, t_{er}^*, t_{ep}^*, t_{sh}^*, t_{sh1}^*$

Table 1: Percent-relative efficiency (PRE) of the various estimators

$PRE(t^*, \bar{y}^*)$	$(1/h)$			
	$(1/2)$	$(1/3)$	$(1/4)$	$(1/5)$
$PRE(t_r^*, \bar{y}^*)$	123.077	105.797	95.35	89.32
$PRE(t_{ks}^*, \bar{y}^*)$	21.62	21.92	22.16	22.60
$PRE(t_{er}^*, \bar{y}^*)$	204.8	182.50	169.95	162.12
$PRE(t_{ep}^*, \bar{y}^*)$	41.83	42.44	43.16	44.072
$PRE(t_{sh}^*, \bar{y}^*)$	210.84	189.22	176.17	167.46
$PRE(t_{sh1}^*, \bar{y}^*)$	212.21	190.71	177.80	169.24

7. CONCLUSION

In this paper, we have proposed twos new estimators (t_{sh}^* & t_{sh1}^*) of the finite population mean of the study variable using auxiliary information. Compared to usual mean and other estimated estimates, the proposed estimation methods are compared. To support the theoretical results, a numerical study is carried out. It should be considered from table 1 that the PRE of t_r^*, t_{er}^*, t_{sh}^*, and t_{sh1}^* decreases while the PRE of t_{ks}^* and t_{ep}^* increases with the value of h increase. Furthermore, it is observed that the estimators t_{sh}^* and t_{sh1}^* are best among $\bar{y}^*, t_r^*, t_{ks}^*, t_{er}^*$ and t_{ep}^* but best among all proposed estimators is t_{sh1}^*. A comparative study was performed and demonstrated that the proposed estimators perform better than the conventional estimator. Therefore, the proposed estimators t_{sh}^* and t_{sh1}^* should be recommended for use in practice.

ACKNOWLEDGMENT

The authors are thankful to the learned referees for their valuable suggestions and comments.

CONFLICT OF INTEREST

No potential conflict of interest was reported by the author(s).

ORCID ID

Udita Gupta https://orcid.org/0009-0000-1754-1418

8. REFERENCES

1. Bahl, S., Tuteja, R.K. (1991). Ratio and Product type exponential estimator. *Infrm. and Optim. Sci.* XII(I): 159-163.
2. Bedi, P.K. (1996). Efficient utilization of auxiliary information at estimation stage, *Biom. J.* 38(8): 973-976.
3. Cochran, W. G. (1977). *Sampling Techniques*, 3rd ed., John Wiley and Sons, New York.
4. Grover, L.K., Kaur, P. (2011). An improved estimator of the finite population mean in simple random sampling. *Model Assisted Stat. Appl.* 6(1):47–55.
5. Grover, L.K., Kaur, P. (2014). A generalized class of ratio type exponential estimators of population mean under linear transformation of auxiliary variable. *Commun. Stat. Simul. Comput.* 43(7):1552–1574.
6. Hansen, M. H. and Hurwitz, W. N. (1946). The problem of non-response in sample surveys. *Journal of the American Statistical Association*, 41: 517–529.
7. Khare, B. B. and Srivastava, S. (1993). Estimation of population mean using auxiliary character in presence of non-response. *The National Academy of Sciences, Letters, India*. 16: 111–114.

8. Khare, B. B. and Srivastava, S. (1995). Study of conventional and alternative two-phase sampling ratio, product and regression estimators in presence of non-response. *Proceedings of the National Academy of Sciences, India*. 65(A): 195–203.
9. Khare, B. B. and Srivastava, S. (1997). Transformed ratio type estimators for the population mean in the presence of non-response. *Communications in Statistics - Theory and Methods.* 26: 1779–1791.
10. Kumar, S. and Bhougal, S. (2011). Estimation of the Population Mean in presence of non-response. *Communication of the Korean Statistical Society.* 18(4): 537-548.
11. Lewisi, P. A., Jones, P. W., Polak, J. W. and Tillotson, H. T. (1991). The problem of conversion in method comparison studies. *Journal of the Royal Statistical Society. Series C (Applied Statistics).*40: 105–112.
12. Okafor, F. C. and Lee, H. (2000). Double sampling for ratio and regression estimation with sub-sampling the non-respondents. *Survey Methodology*. 26: 183–188.
13. Rao, P. S. R. S. (1986). Ratio estimation with sub sampling the non-respondents. *Survey Methodology*.12: 217–230.
14. Rao, P. S. R. S. (1987). Ratio and regression estimate with sub sampling the non-respondents. Paper presented at a 6pecial contributed session of the International Statistical Association Meeting. Sept.: 2–16, Tokyo, Japan.
15. Singh, D. and Choudhary, F. S. (1986). *Theory and Analysis of Sample Survey Designs*. Wiley Eastern Limited. New Delhi. :108.
16. Singh, H. P. and Kumar, S. (2008). A regression approach to the estimation of finite population means in presence of non-response. *Australian and New Zealand Journal of Statistics*. 50: 395–408.
17. Singh, H. P. and Kumar, S. (2009a). A general class of estimators of the population mean in survey sampling using auxiliary information with sub sampling the non-respondents. *The Korean Journal of Applied Statistics.* 22: 387–402.
18. Singh, H. P. and Kumar, S. (2009b). A general procedure of estimating the population mean in the presence of non-response under double sampling using auxiliary information. *SORT*. 33: 71–84.
19. Singh, H. P. and Kumar, S. (2010). Estimation of mean in presence of non-response using two phase sampling scheme. *Statistical Papers*. 50: 559–582.
20. Singh, H. P., Kumar, S. and Kozak, M. (2010). Improved estimation of finite population mean when sub-sampling is employed to deal with non-response. *Communication in Statistics - Theory and Methods*. 39: 791–802.
21. Singh, H. P. and Ruiz Espejo, M. (2003). On linear regression and ratio-product estimation of a finite population mean. *Journal of the Royal Statistical Society. Series D (The Statistician)*. 52: 59–67.
22. Singh, R., Chauhan, P. and Sawan, N. (2008). On linear combination of ratio and product type exponential estimator for estimating the finite population mean. *Statistics in Transition - New Series*. 9: 105–115.
23. Srinath, K. P. (1971). Multiphase sampling in non-response problems. *Journal of the American Statistical Association*. 66: 583–586.
24. Tabasum, R. and Khan, I. A. (2004). Double sampling for ratio estimation with non- response. *Journal of the Indian Society of Agricultural Statistics*. 58: 300–306.
25. Tabasum, R. and Khan, I. A. (2006). Double sampling ratio estimator for the population mean in presence of non-response. *Assam Statistical Review*. 20: 73–83.